Towards process-based narratives for seasonal climate predictions

Johanna Baehr

Mikhail Dobrynin\textsuperscript{2,1}, Nele Charlotte Neddermann\textsuperscript{1}, André Düsterhus\textsuperscript{3}, Kristina Fröhlich\textsuperscript{2}, Wolfgang A. Müller\textsuperscript{4}, Tim Stockdale\textsuperscript{5}, Anca Brookshaw\textsuperscript{5}, Adam Scaife\textsuperscript{6}, Panos Athanasiadis\textsuperscript{7}

\textsuperscript{1} Institute of Oceanography, Center for Earth System Research and Sustainability (CEN), Universität Hamburg, Germany
\textsuperscript{2} Deutscher Wetterdienst (DWD), Germany
\textsuperscript{3} Maynooth University, Ireland
\textsuperscript{4} Max Planck Institute for Meteorology, Hamburg, Germany
\textsuperscript{5} ECMWF, UK
\textsuperscript{6} Hadley Centre, Met Office, UK and College of Engineering, Mathematics and Physical Sciences, University of Exeter, UK
\textsuperscript{7} CMCC - Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, Italy
Setting the scene

Let’s take an ensemble prediction system and investigate seasonal predictive skill for winter and summer over Europe.

More specifically, let’s start with

- ‘winter’: DJF mean (initialized in Nov)
- ‘seasonal predictive skill’: anomaly correlation coefficient (ACC) over re-forecast period
- ‘ensemble prediction system’: 20 years of re-forecasts with multi-model Copernicus C3S ensemble

<table>
<thead>
<tr>
<th>N members</th>
<th>Initialisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMCC</td>
<td>40</td>
</tr>
<tr>
<td>ECMWF</td>
<td>25</td>
</tr>
<tr>
<td>GCFS2</td>
<td>30</td>
</tr>
<tr>
<td>MeteoFrance</td>
<td>15</td>
</tr>
<tr>
<td>MPI-ESM MR</td>
<td>30</td>
</tr>
<tr>
<td>UKMO</td>
<td>28</td>
</tr>
<tr>
<td>Multi-model</td>
<td>138</td>
</tr>
</tbody>
</table>
Anomaly correlation for multi-model mean T2m not particularly high over Europe

Dobrynin et al., in prep.
Temperature anomalies strongly vary over different winters

Temperature anomaly for 2007

Temperature anomaly for 2010

Dobrynin et al., in prep.
... but the ensemble mean neither reproduces spatial structure or amplitude

Dobrynin et al., in prep.
Yet, we hypothesize that the NAO plays an important role

Temperature anomaly for 2007

Temperature anomaly for 2010

Correlation between DJF T2m and DJF NAO in ERAinterim

Dobrynin et al., in prep.
Multi-model NAO: large spread and low correlation

Dobrynin et al., in prep.
Individual model NAO: large spread and a range of correlations

Dobrynin et al., in prep.
Could teleconnections help?

ERAinterim DJF T2m anomaly

ERAinterim Oct SST anomaly

ERAinterim Oct 100hPa Temp anomaly

Dobrynin et al., in prep.
Four predictors consistently suggest NAO+ for 2007 and NAO- for 2010

SST (Oct), Sea ice concentration (Sep), Snow depth (Oct), Stratospheric temperature (Oct)

Dobrynin et al., in prep.
We know only keep ensemble members that are close to at least one predictor:

- SST (Oct)
- Sea ice concentration (Oct)
- Snow depth (Oct)
- Stratospheric temperature (Oct)

First guess DJF NAO

Forecast period (following a training period)

Dobrynin et al., in prep.
The refined multi-model NAO shows higher corr and std than ensemble mean.
The refined individual model NAOs shows higher corr and std than ensemble mean

Dobrynin et al., in prep.
Anomaly correlation for multi-model mean T2m improves over Europe

Dobrynin et al., in prep.
Is this improving in the right direction? Thought experiment
Anomaly correlation for multi-model mean T2m improves in expectable regions, but not everywhere

Dobrynin et al., in prep.
Is this improving in the right direction? Back to T2m anomalies for 2007 and 2010

Dobrynin et al., in prep.
T2m anomalies for 2007 and 2010 improve in pattern and amplitude

Temperature anomaly for 2007

Temperature anomaly for 2010

Dobrynin et al., in prep.
Preliminary summary

Using teleconnections and the state of different predictors immediately before the start for the re-forecast the correlation and signal (amplitude) of both NAO and T2m over Europe can be improved.
... and what about summer?
Little seasonal re-forecast skill for European summer climate

Figure 3: Hindcast skill depicted by Anomaly Correlation (ACC) for 500 hPa geopotential height anomalies (Z500) in summer (JA) comparing the model predictions of MPI-ESM to ERA-20C for different ensemble means (left column), and averaged Z500 values in Europe in the area [35° N, 10° W - 30° E] depicting all ensemble members (grey circles) and selected members (filled grey dots) and their mean (coloured lines) in comparison to ERA-20C (black line), including correlation values (right column). Ensemble means are taken over (a)-(b) all ensemble members in 1902-2008, and cluster selected members in 1930-2008 for (c)-(d) the ensemble members in the predicted dominant cluster, (e)-(f) the ensemble members in the observed cluster, and (g)-(i) the ensemble members in the group of clusters observed. Black dots show significance at the 95% confidence level.

Single Model re-forecasts:
- MPI-ESM-MR
- 1902-2010
- started very May
- 10 members

Compared against 20CR

Neddermann et al., in prep.
Summer NAO only explains about about 30 % of the variance

1. EOF of Z500, 1982-2016, (29%)

North Atlantic Oscillation (NAO) + Zonal Pressure Difference (PD) +
(e.g. Wulff et al., 2017)

2. EOF of Z500, 1982-2016, (19%)

Why does the ensemble spread out?

• different physical mechanisms influence European summers:

→ Is MPI-ESM able to represent those mechanisms & how is skill affected?

Neddermann et al., in prep.
... zonal pressure difference might explain another 20 % of the variance

1. EOF of Z500, 1982-2016, (29%)
   North Atlantic Oscillation (NAO) + Zonal Pressure Difference (PD) +
   (e.g. Wulff et al., 2017)

2. EOF of Z500, 1982-2016, (19%)

Why does the ensemble spread out?
• different physical mechanisms influence European summers:

3 → Is MPI-ESM able to represent those mechanisms & how is skill affected?

(e.g. Folland et al., 2009)

(e.g. Wulff et al., 2017)

Neddermann et al., in prep.
In other words, European summer climate variability influenced by (at least) 4 clusters.

**ERA-20C**

NAO+ (23.1%)

NAO- (33.3%)

PD+ (28.2%)

PD- (16.7%)

Neddermann et al., in prep.
... whose structure and frequency the model ensemble can approximately reproduce

> One cluster assigned to each summer.

Neddermann et al., in prep.
However, the ensemble does not reproduce the ‘timing’

Neddermann et al., in prep.
Idealized thought experiment: correlation skill, if one cluster dominated, which we got right
Idealized thought experiment: correlation skill, if one cluster dominated, which we got right
Yet, this re-forecast skill is highly time-dependent.

Figure 3: Hindcast skill depicted by Anomaly Correlation (ACC) for 500 hPa geopotential height anomalies (Z500) in summer (JA) comparing the model predictions of MPI-ESM to ERA-20C for different ensemble means (left column), and averaged Z500 values in Europe in the area $[35\degree N, 10\degree W, 30\degree E]$ depicting all ensemble members (grey circles) and selected members (filled grey dots) and their mean (coloured lines) in comparison to ERA-20C (black line), including correlation values (right column). Ensemble means are taken over (a)-(b) all ensemble members in 1902-2008, and cluster selected members in 1930-2008 for (c)-(d) the ensemble members in the predicted dominant cluster, (e)-(f) the ensemble members in the observed cluster, and (g)-(i) the ensemble members in the group of clusters observed. Black dots show significance at the 95% confidence level.

Figure 4: Time dependency of the hindcast skill shown by (a) the evolution of the anomaly correlation from 1930-2008 for the mean over all ensemble members (grey), and the mean over all members assigned to the positive NAO (red), the negative NAO (blue), and to the observed cluster in each year (black). (b)-(i) Anomaly Correlation (ACC) for each of those ensemble means averaged in 1930-1970 (left column) and 1970-2008 (right column). Black dots show significance at the 95% confidence level.

Neddermann et al., in prep.
Reality: correlation skill, if one cluster dominated

Figure 2: (left column) Hindcast skill depicted by the Anomaly Correlation Coefficient (ACC) for Z$_{500}$ anomaly in summer (JA) comparing the model predictions of MPI-ESM-MR to ERA-20C for different ensemble means. Black dots show significance at the 95% confidence level. (right column) Averaged Z$_{500}$ anomaly in Europe in the area [35$^\circ$N, 10$^\circ$W, 30$^\circ$E] for all ensemble members (grey circles), selected members (filled grey dots), and their mean (coloured lines) in comparison to ERA-20C (black line), including correlation values.

Ensemble means are taken over:

- Option 1: all ensemble members in all clusters for 1902-2008,
- Option 2: ensemble members selected for the predicted dominant clusters for 1930-2008,
- Option 3: ensemble members selected for the observed clusters for 1930-2008.

Figure 3: Hindcast skill depicted by Anomaly Correlation (ACC) for 500 hPa geopotential height anomalies (Z$_{500}$) in summer (JA) comparing the model predictions of MPI-ESM to ERA-20C for different ensemble means (left column), and averaged Z$_{500}$ values in Europe in the area [35$^\circ$N, 10$^\circ$W, 30$^\circ$E] depicting all ensemble members (grey circles) and selected members (filled grey dots) and their mean (coloured lines) in comparison to ERA-20C (black line), including correlation values (right column).

Ensemble means are taken over:

- Option 1: all ensemble members in 1902-2008, and cluster selected members in 1930-2008 for Option 2: the ensemble members in the predicted dominant cluster, (e)-(f) the ensemble members in the observed cluster, and (g)-(i) the ensemble members in the group of clusters observed. Black dots shows significance at the 95% confidence level.

Neddermann et al., in prep.
Figure 3: Hincast skill depicted by Anomaly Correlation (ACC) for 500 hPa geopotential height anomalies (Z500) in summer (JA) comparing the model predictions of MPI-ESM to ERA-20C for different ensemble means (left column), and averaged Z500 values in Europe in the area [35N, 70N, 10W, 30E] depicting all ensemble members (grey circles) and selected members (filled grey dots) and their mean (coloured lines) in comparison to ERA-20C (black line), including correlation values (right column). Ensemble means are taken over (a)-(b) all ensemble members in 1902-2008, and cluster selected members in 1930-2008 for (c)-(d) the ensemble members in the predicted dominant cluster, (e)-(f) the ensemble members in the observed cluster, and (g)-(i) the ensemble members in the group of clusters observed. Black dots show significance at the 95% confidence level.

Faint hope: are there ways to predict the cluster?

Neddermann et al., in prep.
While anomaly correlation for JA temperature over Europe is small

Re-forecast set

- for May initialization between 1982-2016 with 30 ensemble members each
- with MPI-ESM-MR against ERAinterim

Neddermann et al., 2018
... SSTs show prediction skill 2 – 4 months ahead,

Re-forecast set
- for May initialization between 1982-2016 with 30 ensemble members each
- with MPI-ESM-MR against ERAinterim

Neddermann et al., 2018
... the ensemble spread encompasses the observed state at most times.

Neddermann et al., 2018
Back to teleconnections

- Use a mechanism
- Refine ensemble: remove all members that do not follow the evolution of the entire mechanism
- Recalculate mean
Back to teleconnections

- **Use a mechanism**
- Refine ensemble: remove all members that do not follow the evolution of the **entire** mechanism
- Recalculate mean
Tropical SSTs in April are connected to a wave train in July-August.

Warm SSTs are the source of strong convection in the tropical region, which act as a Rossby wave source (e.g., Gastineau & Frankignoul, 2015), known as circumglobal wave train (e.g., Ding & Wang, 2005).
The wave-train is connected to European summer temperatures

The wave train generates an east-west pressure gradient (Wulff et al, 2017). The zonal pressure gradient in turn is related to the surface temperatures over central Europe (Saeed et al., 2014; Wulff et al, 2017).

Neddermann et al., 2018
Back to teleconnections

- Use a mechanism
- Refine ensemble: remove all members that do not follow the evolution of the entire mechanism

Analyse sign of SST anomaly in April

Cluster analysis of wave train (JA) – retain only members matching the phase proposed by April SST

Analyse zonal SLP difference (JA) – retain only members matching the proposed wave train phase

Analyse mean SAT over central Europe (JA) - retain only members matching the proposed wave train phase

- Recalculate mean
Refined ensemble shows higher correlation skill for central Europe

Fig. 6: Anomaly correlation (ACC) derived without cross-validation for (first row) surface temperatures and derived as the mean over all leave-one-out cross-validated correlation values for (second row) surface temperatures, (third row) SLP and (fourth row) 500 hPa geopotential height (Z500) in summer (JA), comparing the model predictions of MPI-ESM to ERA-Interim in 1982-2016. The ensemble mean is taken over (a)-(d) the full ensemble and (e)-(h) the selected ensemble. Dots show significance at the 95% confidence level, hatching represents areas in which significance is reached in every leave-one-out cross-validated iteration.

Correlation skill for ensemble mean JA SAT

Correlation skill for refined ensemble mean JA SAT

Neddermann et al., 2018
What does the selection actually do?

Refined ensemble shows smaller spread than ensemble mean.

Neddermann et al., 2018
What does the selection actually do?

Refined ensemble shows ‘flatter’ rank histogram than ensemble mean.
How do we know this working in the right direction? Another thought experiment

Skill increase limited to regions, where suggested mechanism is expected to influence variability.

**Fig. 6:** Anomaly correlation (ACC) derived without cross-validation for (first row) surface temperatures and derived as the mean over all leave-one-out cross-validated correlation values for (second row) surface temperatures, (third row) SLP and (forth row) 500 hPa geopotential height (Z500) in summer (JA), comparing the model predictions of MPI-ESM to ERA-Interim in 1982-2016. The ensemble mean is taken over (a)-(d) the full ensemble and (e)-(h) the selected ensemble. Dots shows significance at the 95% confidence level, hatching represents areas in which significance is reached in every leave-one-out cross-validated iteration.

**Fig. 9:** Hindcast skill for the “perfect” ensemble selection derived with the known state of the SLP index. (a)-(b) Anomaly correlation (ACC) for (a) surface temperatures and (b) 500 hPa geopotential height (Z500) in summer (JA), comparing the model predictions of MPI-ESM to ERA-Interim in 1982-2016 derived as the mean over all leave-one-out cross-validated correlation values. Hatching represents areas in which significance is reached in every leave-one-out cross-validated iteration. (c)-(d) Reliability diagrams comparing the mean over the selected (red) and the “perfect” (blue) ensemble of MPI-ESM to ERA-Interim in 1982-2016 for (c) temperature and (d) Z500 in the area [35°N, 20°W]. Vertical lines show the error bootstrapped at the 95% confidence level. The histograms depict the distribution of the data.

*Neddermann et al., in review*
What about the NAO?

EOF #2 JA SLP

Correlation of PC#2 JA SLP with SAT

Correlation skill for ‘perfect’ ensemble JA SAT

EOF #1 JA SLP

Correlation of PC#1 JA SLP with SAT

Neddermann et al., 2018
What about the NAO?

EOF #2 JA SLP

Correlation of PC#2 JA SLP with SAT

Correlation skill for ‘perfect’ ensemble JA SAT

Limited skill increase for regions strongly influenced by NAO.

EOF #1 JA SLP

Correlation of PC#1 JA SLP with SAT

Neddermann et al., 2018
Preliminary summary

Using teleconnections and the state of different predictors immediately before the start for the re-forecast ...

... can improve correlation and signal (amplitude) of surface temperature over Europe for seasonal winter and summer re-forecasts.
Final remarks

An ensemble analysis guided by a process-based narrative

- starts from the initial conditions and tells a physically consistent story by only retaining those ensemble members that follow,
- has to be tailored to every process/region and season,
- still makes use of the strengths of a dynamical ensemble,
- shifts focus from the ensemble-mean and re-forecast skill mean closer to individual years.

When the ensemble spread is large, the ensemble resulting from a process-based narrative

- looks considerably different than the ensemble mean,
- shows increased ‘signal’ and increased skill,
- can provide insights into model performance and deficits.